Title: Is there a clinically meaningful difference in patient reported dyspnea in acute heart failure? An analysis from URGENT Dyspnea

Running title: Minimal clinical important difference of dyspnea in AHF

Authors: Peter S. Pang MD, MSc,[1] Kathleen A. Lane MS,[2] Miguel Tavares MD PhD,[3] Alan B. Storrow MD,[4] Changyu Shen PhD,[2] W. Frank Peacock MD,[5] Richard Nowak MD,[6] Alexandre Mebazaa MD, PhD,[7] Said Laribi MD, PhD,[8] Judd E. Hollander MD,[9] Mihai Gheorghiade MD,[10] Sean P. Collins MD, MSc.[4]

Author affiliations:

[1]Department of Emergency Medicine, Indiana University School of Medicine

[2]Department of Biostatistics, Indiana University School of Medicine

[3]Department of Anesthesiology and Critical Care, Hospital Geral de Santo António, Porto, Portugal

[4]Department of Emergency Medicine, Vanderbilt University and the Nashville VA

[5]Department of Emergency Medicine, Baylor College of Medicine, Houston, Texas

[6]Department of Emergency Medicine, Henry Ford Health System, Wayne State University

[7]Department of Anesthesia and Critical Care, Hôpitaux Universitaires Saint Louis Lariboisière

This is the author's manuscript of the article published in final edited form as:
Pang, P. S., Lane, K. A., Tavares, M., Storrow, A. B., Shen, C., Peacock, W. F., ... & Gheorghiade, M. (2017). Is there a clinically meaningful difference in patient reported dyspnea in acute heart failure? An analysis from URGENT Dyspnea. Heart & Lung: The Journal of Acute and Critical Care. https://doi.org/10.1016/j.hrtlng.2017.03.003

[8]Tours University Hospital, Emergency Department, 37044, and INSERM, U942, BIOmarkers in CArdioNeuroVAScular diseases

[9]Sidney Kimmel Medical College of Thomas Jefferson University

[10]Division of Cardiology, Northwestern University Feinberg School of Medicine

Corresponding Author:

Peter S. Pang MD MSc FACEP FAAEM FAHA FACC

Associate Professor | Emergency Medicine

Associate Director | Clinical Research

Department of Emergency Medicine

Indiana University Translational Sciences Fellow

Indiana University School of Medicine

ppang@iu.edu

317-880-8900

Manuscript Word Count = 3517 (including references)

Abstract

Background: Dyspnea is the most common presenting symptom in patients with acute heart failure (AHF). However, the minimal clinically important difference (MCID) to patients has not been well established.

Methods: We performed a secondary analysis from URGENT Dyspnea, an observational, multi-center study of AHF patients enrolled within one hour of first physician assessment in the ED. Three scales were used to assess dyspnea: 1) 10cm VAS 2) 5-point Likert, and 3) a 7-point Likert (both VAS and 5-point Likert were recorded in the upright and supine positions). Using the anchor-based method to determine the MCID, a one-category change in the 7-point Likert was used as the criterion standard ('minimally improved or worse').

Results: Of the 776 patients enrolled, 491 had a final diagnosis of AHF with responses at both time points. A 10.5mm (SD 1.6 mm) change in VAS was the MCID for improvement in the upright position, and 14.5mm (SD 2.0mm) in the supine position. However, there was no MCID for worsening, as few patients reported worse dyspnea. There was also no significant MCID for the 5-point Likert scale.

Conclusion: A 10.5 mm change is the MCID for improvement in dyspnea over 6 hours in ED patients with AHF.

Key Words: dyspnea, acute heart failure, emergency department, MCID

Introduction

Patient reported outcomes (PRO) are defined as 'any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patients response by anyone else.'[1] As a measurement of patients' experiences, PRO are key assessments in patient centered research. Dyspnea, or the sensation of breathlessness, is one of the most commonly measured PRO's in acute heart failure (AHF) clinical trials.

The sensation of difficulty breathing or shortness of breath compels patients with AHF to seek medical care.[2,3] Early and persistent relief of dyspnea has been associated with improved outcomes.[4-7] Although dyspnea is significantly improved after initial therapy,[8] a substantial number of patients continue to have shortness of breath during hospitalization.[4-6] As such, its relief is important to both patients and caregivers, especially with the current focus on patient centered outcomes.[9]

As a subjective, patient reported symptom, how exactly to assess and measure dyspnea continues to be debated.[9-11] While clinical trials now use a more standardized method of dyspnea assessment — formal training, standardized position, only after a period of rest[7] — use of dyspnea as a clinical trial endpoint has fallen out of favor, in part due to the difficulty of demonstrating a significant difference between investigational agents and usual care.[12,13] However, as the predominant AHF symptom, relief from dyspnea is important to patients. Similar to the measurement of pain, proper measurement of dyspnea in AHF is needed.[9]

The minimal clinically important difference (MCID) is the “smallest benefit of value to patients.” [14] As clinicians and patients may disagree on what is clinically meaningful, understanding patients’ perspective is critical for a patient centered outcome. Knowing the MCID also informs clinical trial design, providing the minimal effect size. Despite the importance of dyspnea to patients and its near universal presence in AHF patients, the minimal clinically important difference (MCID) in dyspnea via various measurement scales has not been well studied.[7] Thus, we performed a retrospective analysis of the URGENT-Dyspnea database to determine the MCID in dyspnea at 6 hours after initial management in patients with AHF presenting to the emergency department (ED).

Methods:

Details regarding the URGENT Dyspnea (Ularitide Global Evaluation in Acute Decompensated Heart Failure) study design and main study results have been previously presented.[8] Briefly, URGENT Dyspnea was a multi-center, prospective observational study that enrolled 776 patients from 17 countries involving 35 sites from January through August of 2007. The primary objective was to determine changes in patient reported dyspnea over 6 hours, capturing patients shortly after ED presentation. The study was IRB or Ethics Committee approved at every study site.

Participants

Patients were 18 years and older and enrolled within one hour of first physician contact. Given the short time frame, patients with dyspnea presumed attributable to AHF were approached, consented, and then enrolled. To best replicate 'real-world' conditions, inclusion and exclusion criteria were intentionally kept broad. Treatment and management were directed by the patients' clinical care team: there were no pre-specified protocols or treatment interventions. Demographic, clinical, and treatment data were collected per standardized case report form. The site principal investigator, who had full access to all available clinical data, determined the final diagnosis of AHF.

Dyspnea Assessment Instruments

At 6 hours after enrollment, patients were asked about the severity of their dyspnea. They were asked to report via commonly used scales in AHF. The 7-point Likert scale: “Compared to how you felt when you first arrived, do you now feel your breathing is: *Markedly worse, moderately worse, minimally worse, no change, minimally improved, moderately improved, markedly improved?”* A one-category change of “minimally worse” or “minimally improved” was used as the criterion standard for the MCID in this study. This standard was chosen based on previously published work in AHF and the MCID,[15] which was based on prior work in the assessment of pain.[15-17]

Two other scales were used to assess dyspnea at both time zero and 6 hours later; a 5-point Likert scale (“I am not short of breath (SOB)”, “Mildly SOB”, “Moderately SOB”, “Severely SOB”, “Very Severely SOB”) and a 100mm VAS, with 0 as “I am not breathless at all” to 100mm as “I am the most breathless I have ever been.” Per protocol, this 100mm line was divided into 10 equal one centimeter increments and scored accordingly.

Outcome Measures

The main outcome measures were the change in visual analog scale (VAS) and 5-point Likert scale from baseline to 6-hour assessment relative to a 1-category change response in the 7-point Likert scale (‘minimally worse’, ‘no change’, or ‘minimally better’).

Both the 5-point Likert and VAS were measured in two positions to ascertain the effect of position on patient reported dyspnea. Patients were

initially assessed in the upright position (seated, head of bed ≥ 60 degrees). If patients reported “severely” or “very severely” by the 5-point Likert, the supine position was not assessed due to safety concerns. For any other score, patients were placed in the supine position (head of bed ≤20 degrees) and after an equilibration period of 120seconds, both the 5-point Likert and 100mm VAS were repeated.

Analysis Plan

We utilized the anchor based method to determine the MCID.[14] To the best of our knowledge, the only other study to explore the MCID in AHF from the ED perspective also utilized the anchor based method.[15] This method uses another measure of improvement – the 7-point Likert scale – as the ‘anchor’ to associate change via another numerical scale.[14]

As we were not certain what the MCID would be, we did not pre-specify the effect size. However, using conservative estimates, we have 80% power to detect a 10mm change in VAS corresponding to a 1 point minimal improvement by Likert, assuming a group size of 64 subjects and type 1 error controlled at 0.05 (two-sided)

Statistical Analysis

Patients were divided into 3 groups based upon the 7-point Likert; 1) those reporting “No Change”, 2) those reporting “Minimally worse”, and 3) those reporting “Minimally improved”. Using this grouping, all other categories were

excluded. To ascertain whether greater or lesser changes had value, additional analyses were performed by dividing the patients into three groups based on any reported improvement, any worsening, or no change in the 7-point Likert, thus using all eligible subjects.

Baseline characteristics including frequencies and percentages for categorical variables and means and standard deviations for continuous variables, were calculated and compared between the groups using Fisher's exact tests and one-way analysis of variance (ANOVA) models, as appropriate.

Change scores between responses at baseline and 6 hours on the 5-point Likert scale items and the VAS were calculated by subtracting the baseline value from the 6-hour value. These change scores were categorized as improvement, worsening and no change. Fisher's exact tests were used to compare the 3 groups by either definition on changes in the 5-point Likert scale and on the VAS in both the upright and supine positions. Separate one-way analysis of variance (ANOVA) models were used to determine if there were any significant differences between the groups based on 6-hour changes for the 5-point Likert and VAS in both the upright and supine positions. For ease in interpretation, the change scores were reverse coded so that a positive score indicated improvement. Following a significant group effect, unadjusted pairwise comparisons were made between the groups. Least square means and standard errors were reported from the models. Kruskal Wallis tests were also used to look for overall group differences; as the results agreed with the ANOVA results, they were not presented. To determine concordance between the anchor scale and the 5-point

Likert and VAS, a kappa (**κ**) statistic was calculated and presented in the tables with 95% CI.

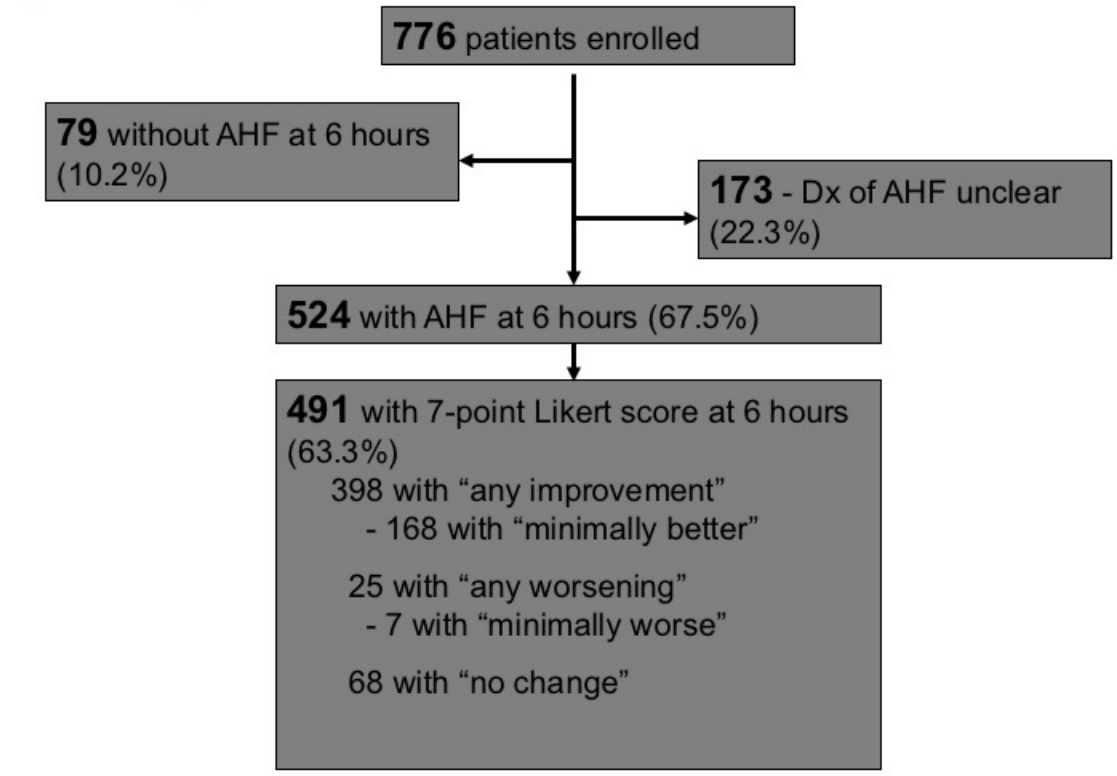
We did not control for multiple-comparisons for the analysis performed in this manuscript and have listed this in the limitations. All analyses were performed using SAS, version 9.4 (SAS, Inc., Cary, NC).

Results

Baseline Characteristics

Figure 1 demonstrates the derivation of the final patient subset for analysis.

Figure 1: Study Patient Flow



Of the 776 patients enrolled, 491 had both a final diagnosis of AHF and baseline/6 hour self-reported dyspnea; 93 reported 'minimally improved' and 7 reported 'minimally worse' by 7-point Likert scale 6 hours after enrollment. Table 1 shows the baseline characteristics of patients grouped by response on the 7-point Likert.

Table 1 – Baseline Characteristics Grouped by Patient-Reported 'Minimally Worse', 'No Change', or 'Minimally Improved' by 7-Point Likert Scale

	Minimally Improved (n=93)	*No Change (n=68)*	*Minimally Worse (n=7)*	*3-group P-value*	*Minimally Improved vs. No Change P-value*
Age, mean ± SD	68.3 ± 14.7	62.6 ± 16.4	62.0 ± 11.0	0.0575	0.0228
Male gender, n (%)	52 (55.9%)	42 (61.8%)	4 (57.1%)	0.7931	0.5185
White race, n (%)	74 (80.4%) (n=92)	50 (73.5%)	4 (57.1%)	0.2148	0.3410
LVEF, mean ± SD	39.7 ± 18.1 (n=59)	36.6 ± 16.2 (n=43)	41.0 ± 13.5 (n=3)	0.6559	0.3813
HISTORY, n (%)					
Heart Failure	(n=91)			0.9491	0.8503
History of heart failure	69 (75.8%)	53 (77.9%)	6 (85.7%)	.	.
de novo heart failure	22 (24.2%)	15 (22.1%)	1 (14.3%)	.	.
Coronary artery disease	50 (53.8%)	26 (38.2%)	3 (42.9%)	0.1484	0.0568
Prior myocardial infarction	20 (21.5%)	18 (26.5%)	2 (28.6%)	0.6882	0.5734
Valvular disease	14 (15.1%)	14 (20.6%)	0 (0.0%)	0.4040	0.4035
Primary Cardiomyopathy	19 (20.4%)	18 (26.5%)	1 (14.3%)	0.6425	0.4487
Prior percutaneous coronary intervention (PCI)	10 (10.8%)	6 (8.8%)	3 (42.9%)	0.0555	0.7932

Prior coronary bypass grafts	11 (11.8%)	5 (7.4%)	0 (0.0%)	0.5658	0.4300
Stroke (CVA)	8 (8.6%)	4 (5.9%)	0 (0.0%)	0.7419	0.5621
Obesity	34 (36.6%)	12 (17.6%)	4 (57.1%)	0.0079	0.0129
Peripheral vascular disease	7 (7.5%)	1 (1.5%)	0 (0.0%)	0.1975	0.1398
Asthma/COPD	17 (18.3%)	8 (11.8%)	1 (14.3%)	0.5256	0.2808
Diabetes - Insulin Dependent	15 (16.1%)	13 (19.1%)	2 (28.6%)	0.5667	0.6764
Diabetes - Non-Insulin Dependent	16 (17.2%)	13 (19.1%)	1 (14.3%)	0.9361	0.8364
Renal insufficiency	26 (28.0%)	18 (26.5%)	2 (28.6%)	0.9545	0.8599
Anemia (HG<12 g/dl)	18 (19.4%)	8 (11.8%)	0 (0.0%)	0.2920	0.2783
Cancer history	6 (6.5%)	4 (5.9%)	1 (14.3%)	0.5097	1.0000
Hypertension	66 (71.0%)	49 (72.1%)	5 (71.4%)	1.0000	1.0000
VITAL SIGNS, mean ± SD					
Systolic Blood pressure (mm Hg)	138.0 ± 34.0	132.9 ± 32.5	127.1 ± 18.5	0.4983	0.3417
Diastolic Blood pressure (mm Hg)	80.0 ± 17.4	76.6 ± 20.8	78.0 ± 15.2	0.5239	0.2607
Heart rate (beats/min)	89.0 ± 25.3	89.4 ± 21.5	82.0 ± 17.9	0.7301	0.9191
Respiratory rate (breaths/min)	22.3 ± 4.9 (n=92)	21.9 ± 4.9 (n=66)	21.3 ± 4.9	0.7767	0.5813
Oxygen saturation (SpO2)	93.5 ± 4.9 (n=90)	95.6 ± 3.6 (n=66)	96.6 ± 2.4	0.0071	0.0030
BASELINE MEDICATIONS, n(%)					
Beta-blockers	56 (60.2%)	41 (60.3%)	5 (71.4%)	0.9246	1.0000
ACE Inhibitors	48 (51.6%)	41 (60.3%)	4 (57.1%)	0.5378	0.3359
Angiotensin receptor blockers	10 (10.8%)	5 (7.4%)	1 (14.3%)	0.4968	0.5871
Statins	33 (35.5%)	20 (29.4%)	0 (0.0%)	0.1241	0.4978
Amlodipine/Felodipine (other dihydropyridine)	11 (11.8%)	8 (11.8%)	0 (0.0%)	1.0000	1.0000

Nitrates	23 (24.7%)	10 (14.7%)	1 (14.3%)	0.3040	0.1661
Aldosterone Antagonists	10 (10.8%)	15 (22.1%)	0 (0.0%)	0.0942	0.0765
Diuretics	70 (75.3%)	50 (73.5%)	5 (71.4%)	0.9537	0.8556
Cardiac Glycoside	20 (21.5%)	14 (20.6%)	0 (0.0%)	0.5866	1.0000
Aspirin	37 (39.8%)	31 (45.6%)	2 (28.6%)	0.6020	0.5194
Plavix	6 (6.5%)	9 (13.2%)	2 (28.6%)	0.0753	0.1743
Coumadin	22 (23.7%)	16 (23.5%)	1 (14.3%)	1.0000	1.0000
Other anti-coagulant	6 (6.5%)	4 (5.9%)	1 (14.3%)	0.5097	1.0000
Pacemaker	7 (7.5%)	4 (5.9%)	2 (28.6%)	0.1591	0.7613
Implantable cardioverter-defribrillators (AICD)	6 (6.5%)	5 (7.4%)	0 (0.0%)	1.0000	1.0000
Cardiac resynchronization therapy (CRT)	2 (2.2%)	0 (0.0%)	0 (0.0%)	0.5492	0.5090
LABORATORY VALUES, mean ± SD					
BNP	1457.8 ±1434.3 (n=33)	1031.2 ±1217.2 (n=26)	620.5 ± 419.5 (n=4)	0.3007	0.2309
Serum sodium (mmol/L)	137.7 ± 4.6 (n=88)	137.8 ± 5.3 (n=66)	139.9 ± 2.4	0.5238	0.8866
Troponin I >0.04 (ng/mL) label	1.9 ± 4.6 (n=29)	2.2 ± 8.2 (n=22)	0.1 ± 0.0 0.2 (n=2)	0.8942	0.8631
Creatinine	1.5 ± 0.7 (n=89)	1.7 ± 1.9 (n=67)	1.2 ± 0.3	0.6119	0.5202
BUN (mg/dL)	35.4 ± 36.8 (n=79)	29.4 ± 24.2 (n=62)	36.9 ± 32.7 (n=6)	0.5107	0.2420
TREATMENT IN THE ED, n (%)					
Loop diuretic	89 (96.7%) (n=92)	57 (86.4%) (n=66)	6 (85.7%)	0.0345	0.0290
Vasodilator	24 (27.6%)	9 (14.1%)	0 (0.0%)	0.0547	0.0719

	(n=87)	(n=64)			
IV Inotrope and/or vasopressor	11 (12.4%) (n=89)	6 (9.2%) (n=65)	1 (14.3%)	0.6803	0.6106
DYSPNEA SCALES					
5 point Likert (upright; n (%))	(n=92)	(n=67)		0.0067	0.0016
I am not SOB	3 (3.3%)	16 (23.9%)	1 (14.3%)	.	
Mildly SOB	36 (39.1%)	23 (34.3%)	2 (28.6%)	.	
Moderately SOB	29 (31.5%)	12 (17.9%)	2 (28.6%)	.	
Severely SOB	18 (19.6%)	11 (16.4%)	1 (14.3%)	.	
Very Severely SOB	6 (6.5%)	5 (7.5%)	1 (14.3%)	.	
5 point Likert (supine, n(%))	(n=66)	(n=48)	(n=4)	<.0001	<0.0001
I am not SOB	1 (1.5%)	13 (27.1%)	1 (25.0%)	.	
Mildly SOB	16 (24.2%)	7 (14.6%)	1 (25.0%)	.	
Moderately SOB	34 (51.5%)	25 (52.1%)	0 (0.0%)	.	
Severely SOB	15 (22.7%)	2 (4.2%)	2 (50.0%)	.	
Very Severely SOB	0 (0.0%)	1 (2.1%)	0 (0.0%)	.	
Visual Analog Scale (upright), mean ± SD	5.2 ± 2.7	4.3 ± 3.0	4.4 ± 3.2	0.0961	0.0323
Visual Analog Scale (supine), mean ± SD	5.7 ± 2.2 (n=67)	4.0 ± 2.3 (n=49)	4.3 ± 3.0 (n=4)	0.0003	<0.0001

Overall, very few patients (n=7) reported feeling ‘minimally worse’ at 6 hours, limiting any comparisons. Characteristics were largely similar between those who reported ‘no change’ to those who reported ‘minimally improved.’ The largest differences were in baseline 5-point Likert responses. Those who reported ‘minimally improved’ were more likely to have worse dyspnea at baseline. (p=.0067) Other differences are as follows: More patients in the

improved category were obese, had lower oxygen saturation at baseline, and had more loop diuretic use.($p < .05$ for all) Supplemental Table 1 shows the characteristics for patients with any improvement or any worsening by 7-point Likert. Similar to the 'minimally improved or worse' group, few patients (n=25) reported feeling worse at 6 hours.

Anchor Scale Based Changes

Table 2 shows the proportion of patients at 6 hours who reported 'minimally worse,' 'no change,' and 'minimally better' scores by 7-point Likert relative to the categorized 6-hour change in the VAS and 5-point Likert scales in both the upright and supine positions.

Table 2: Changes in Likert and VAS Relative to Anchor Scale at 6 hours

	Change in 5-point Likert (Upright) (n=158; p-value=0.0005)			
7-Point Likert at 6 hours	**Improved**	**No Change**	**Worse**	**Total**
Minimally Improved	33 (37.5%)	51 (58.0%)	4 (4.5%)	88
No Change	7 (10.9%)	51 (79.7%)	6 (9.4%)	64
Minimally Worse	3 (50.0%)	2 (33.3%)	1 (16.7%)	6
	Change in VAS (Upright) (n=165; p-value <0.0001)			
Minimally Improved	61 (65.6%)	20 (21.5%)	12 (12.9%)	93
No Change	12 (18.5%)	42 (64.6%)	11 (16.9%)	65
Minimally Worse	2 (28.6%)	0 (0.0%)	5 (71.4%)	7
	Change in 5-point Likert (Supine) (n=112; p-value=0.0067)			
Minimally Improved	27 (42.9%)	34 (54.0%)	2 (3.2%)	63

No Change	8 (17.8%)	31 (68.9%)	6 (13.3%)	45
Minimally Worse	0 (0.0%)	3 (75.0%)	1 (25.0%)	4
	Change in VAS (Supine) (n=116; p-value <0.0001)			
Minimally Improved	48 (73.8%)	14 (21.5%)	3 (4.6%)	65
No Change	16 (34.0%)	22 (46.8%)	9 (19.1%)	47
Minimally Worse	0 (0.0%)	2 (50.0%)	2 (50.0%)	4

Although there are significant differences between groups ($p<0.01$ for all), a sizable proportion of patients in every response category demonstrated discordant results. Patients who reported 'minimally worse' may have reported 'improved' by an alternate scale.

Table 3 shows the proportion of patients who reported any improvement, no change, and any worsening at 6 hours by 7-point Likert and the corresponding frequencies of patients' response by 5-point and VAS.

Table 3: Changes in Likert and VAS Relative to Any Improvement, No Change, or Any Worsening by Anchor Scale at 6 hours

	5-point Likert (Upright) Change (n=467; p-value <0.0001, κ = 0.35, 95% CI (0.27-0.42))			
7-Point Likert at 6 hours	**Improved**	**No Change**	**Worse**	**Total**
Improved	275 (72.6%)	94 (24.8%)	10 (2.6%)	379
No Change	7 (10.9%)	51 (79.7%)	6 (9.4%)	64
Worse	11 (45.8%)	7 (29.2%)	6 (25.0%)	24
	5-point Likert (Supine) Change (n=282; p-value <0.0001. κ = 0.25, 95% CI (0.17-0.35))			
Improved	142 (63.4%)	77 (34.4%)	5 (2.2%)	224
No Change	8 (17.8%)	31 (68.9%)	6 (13.3%)	45
Worse	3 (23.1%)	6 (46.2%)	4 (30.8%)	13
	VAS (Upright) Change (n=485; p-value <0.0001, κ = 0.44, 95% CI (0.36-0.53))			

Improved	331 (83.8%)	44 (11.1%)	20 (5.1%)	395
No Change	12 (18.5%)	42 (64.6%)	11 (16.9%)	65
Worse	14 (56.0%)	0 (0.0%)	11 (44.0%)	25
	VAS (Supine) Change (n=289; p value <0.0001, κ = 0.35, 95% CI (0.24-0.45))			
Improved	187 (81.7%)	29 (12.7%)	13 (5.7%)	229
No Change	16 (34.0%)	22 (46.8%)	9 (19.1%)	47
Worse	4 (30.8%)	4 (30.8%)	5 (38.5%)	13

For patients who reported any improvement by 7-point Likert, they were more likely to report feeling improved by alternate scales. However, the results were inconsistent in regards to reporting 'minimally worse': many patients reported feeling improved by the alternate scale. Overall, there was poor concordance between the scales.

Minimally Clinical Important Difference

The 6-hour change in 5-point Likert scale and VAS, relative to our criterion standard of the MCID — a one-category change in the 7-point Likert scale at 6 hours — are reported in Table 4.

Table 4: Patient Reported Changes in Dyspnea from Baseline to 6 hours on VAS or 5 Point Likert by Anchor Scale – Both One Category Change and Any Change

	VAS Change (mm) UPRIGHT	**P-value**	**VAS Change (mm) SUPINE**	**P-value**
Minimally Improved	10.5 (1.6)		14.5 (2.0)	
No Change	0.3 (1.9)		4.6 (2.3)	
Minimally Worse	-2.9 (5.7)		-12.5 (8.0)	
		<0.0001*		0.0002
Any Improvement	24.7 (1.1)		20.2 (1.3)	
No Change	0.3 (2.7)		4.7 (2.8)	
Any Worsening	10.8 (4.3)		-0.8 (5.3)	
		<0.0001		<0.0001*

	5-point Likert Change UPRIGHT	P-value	5-point Likert Change SUPINE	P-value
Minimally Improved	0.38(0.07)		0.48 (0.09)	
No Change	0.03 (0.09)		0.04 (0.10)	
Minimally Worse	0.67 (0.29)		-0.25 (0.34)	
		0.0045**		0.0021**
Any Improvement	1.12 (0.05)		0.79 (0.05)	
No Change	0.03 (0.13)		0.04 (0.12)	
Any Worsening	0.67 (0.21)		-0.15 (0.22)	
		<0.0001		<0.0001*

Results represent Least Square Means (Standard Errors) from One-Way Analysis of Variance Models on

*No significant difference between No Change and Minimally Worse or Worse. All other pairwise comparisons are significantly different.

** Only the Improved vs. No change category is significantly different. No other pairwise comparisons are significantly different.

Although a significant MCID for improvement was observed, the MCID for worsening was not statistically significant. Changes in position seem to amplify the difference for improvement. A 10.5mm (SE 20mm) change in VAS was the minimal clinical significant improvement difference in the upright position, but 14.5mm (SE 20mm) was in the supine position.

When patients who reported mild, moderate, or marked improvement or worsening were categorized into 'any improvement, no change, or any worsening,' the associated change in VAS was 24.7mm (SD 1.1) in the upright position and 20.2mm (SD 1.3) in the supine positions. For the upright position only, a change of 10.8mm (SD 4.3) in VAS was associated with any worsening. (Table 4) In addition, Table 4 also reports the change relative to the 5-point Likert scale. Although some significant differences were seen, given the 5-point scale

was an ordinal scale, no difference reached the 1-point threshold, with the exception of 'any improvement.'

Of note, to account for the multiplicity of testing, we used the false discover rate (FDR) to determine which values were significant at FDR = 0.05.[18] Based on this criterion, all 16 p-values reported remain significant.

Discussion

In this secondary analysis from the URGENT-Dyspnea registry, a minimal clinical important improvement in dyspnea was slightly greater than 10mm by VAS — in both the upright and supine positions at 6 hours after initial assessment. Of note, the MCID for VAS in the supine position was greater (14.5mm) compared to VAS in the upright position (10.5mm). We have previously reported the effect of positioning on dyspnea response, observing that supine positioning more robustly captures the symptom. [8]No significant MCID in terms of worsening was observed, however. This was driven primarily by the lack of patients who felt worse.

We did not find the MCID for the 5-point Likert scale however, irrespective of position. Although easier to administer, the Likert scale may measure dyspnea differently[8 10] Alternatively, the categorical responses may not discriminate dyspnea sufficiently. We, along with others, have reported differences between the scales despite the same population.[7,8]

To the best of our knowledge, the only other study to explore the MCID in AHF was conducted in ED patients by Ander et.al.[15] Using a similar methodology

of a 1-point difference in a 5-point VAS as the criterion standard for the MCID, Ander et.al. enrolled 74 patients from a single center, and found 21.1mm to be the MCID for both a 'little less difficulty breathing' and a 'little more difficulty breathing.'[15] Patients were assessed every 20 minutes to a maximum of 2 hours, unlike our study, which did a single assessment at 6 hours. Differences between this study and our analysis may have been driven by the frequency of assessments, the short timeframe over which to assess the MCID, geographical variation, or baseline differences in dyspnea severity. We had previously shown that dyspnea improves rapidly within the first 6 hours of therapy.[8]

This method of utilizing one scale to determine the MCID in another scale — anchor based method — has been previously used in AHF[15] as well as pain scales.[16] In fact, significant differences in VAS pain perception, 13mm on a 100mm scale,[17] are remarkably similar to our findings. This further supports the importance of our findings and supports patient reported dyspnea as an appropriate physiologic endpoint in AHF studies, commensurate to that of the experience of pain. Although attempts to correlate a subjective response with objective criteria are understandable,[11] the patients' response continues to be an important perspective. For a symptom such as dyspnea, the patients' response may be the only valid one.[19] However, the MCID may differ based on the phenotype of AHF studied, the timing of assessment, and the setting in which it is assessed.[19] Thus, continued work is needed to better understand the MCID in AHF.

Unlike the MCID, when ‘any worsening’ or ‘any improvement’ was used as the criterion standard, a much greater change in VAS was noted. Interestingly, ‘any worsening’ was associated with improvement by VAS. This finding was also observed with the 5-point Likert. For both scales, this paradoxical finding was attenuated by position. Although there was still improvement, a smaller proportion reported improvement in supine vs. upright position. Overall, the numbers of patients who reported worse dyspnea was small. Another related finding was the lack of concordance between the scales. As discussed in a prior analysis,[10] such discordance suggests each scale may capture different aspects of this subjective symptom.

Enthusiasm for dyspnea as an endpoint has waned in clinical trials.[12] In part, this is due to the difficulty of achieving a substantial effect over usual therapy. However, failure to achieve a differential effect with novel therapies does not mitigate its importance to patients. Additionally, failure to identify a MCID may have also contributed to the waning interest in dyspnea as a trial endpoint. In other fields, such as asthma, where dyspnea is also a predominant symptom, multiple scales and MCID’s are well established. As a result, dyspnea remains a key endpoint.[20] Alternatively, in COPD, although MCID for dyspnea in the chronic setting exist, there is actually no MCID for COPD exacerbations.[21] Future work to confirm our findings are needed.

Limitations

Our study has several limitations. As a secondary analysis, unmeasured confounders may significantly impact our findings. The total number of patients who reported minimal improvement or worsening was also small. Larger numbers may have yielded different results or more narrow confidence intervals. However, our study is one of the larger studies conducted in the ED setting to examine the MCID. In addition, other MCID's have been driven largely by improvement.[22] Our study also uses only a single anchor based method to determine the MCID. Although the anchor is well established in AHF clinical trials, [13] we did not specifically address other key domains, such as construct and content validity. More frequent measurements may have also yielded different results. We also did not control for multiple comparisons, as we considered each comparison to be sufficiently unique so that family-wise type I error is not particularly relevant.

Conclusion

Dyspnea is the most common symptom in patients presenting with AHF. A 10.5 mm change is the MCID for a 6-hour dyspnea improvement in ED patients with AHF. As a patient reported outcome, a better understanding of the MCID may inform future studies targeting this symptom.

Figure Legends:

Figure 1: Study Patient Flow

AHF = Acute Heart Failure

Acknowledgements: None

Funding sources: The original URGENT Dyspnea study was funded by PDL Biopharma. However, there was no support or funding for this analysis.

1. USDHHSFDA. Guidance for Industry: Patient Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. United States2009.
2. Fonarow GC. The Acute Decompensated Heart Failure National Registry (ADHERE): opportunities to improve care of patients hospitalized with acute decompensated heart failure. *Rev Cardiovasc Med.* 2003;4 Suppl 7:S21-30.
3. Fonarow GC, Stough WG, Abraham WT, et al. Characteristics, treatments, and outcomes of patients with preserved systolic function hospitalized for heart failure: a report from the OPTIMIZE-HF Registry. *J Am Coll Cardiol.* 2007;50(8):768-777.
4. Metra M, O'Connor CM, Davison BA, et al. Early dyspnoea relief in acute heart failure: prevalence, association with mortality, and effect of rolofylline in the PROTECT Study. *Eur Heart J.* 2011;32(12):1519-1534.
5. Metra M, Teerlink JR, Felker GM, et al. Dyspnoea and worsening heart failure in patients with acute heart failure: results from the Pre-RELAX-AHF study. *Eur J Heart Fail.* 2010;12(10):1130-1139.
6. Metra M, Cleland JG, Weatherley BD, et al. Dyspnoea in patients with acute heart failure: an analysis of its clinical course, determinants, and relationship to 60-day outcomes in the PROTECT pilot study. *Eur J Heart Fail.* 2010;12(5):499-507.
7. Teerlink JR, Cotter G, Davison BA, et al. Serelaxin, recombinant human relaxin-2, for treatment of acute heart failure (RELAX-AHF): a randomised, placebo-controlled trial. *Lancet.* 2013;381(9860):29-39.
8. Mebazaa A, Pang PS, Tavares M, et al. The impact of early standard therapy on dyspnoea in patients with acute heart failure: the URGENT-dyspnoea study. *Eur Heart J.* 2010;31(7):832-841.
9. Pang PS, Cleland JG, Teerlink JR, et al. A proposal to standardize dyspnoea measurement in clinical trials of acute heart failure syndromes: the need for a uniform approach. *Eur Heart J.* 2008;29(6):816-824.
10. Pang PS, Collins SP, Sauser K, et al. Assessment of dyspnea early in acute heart failure: patient characteristics and response differences between likert and visual analog scales. *Acad Emerg Med.* 2014;21(6):659-666.
11. Ezekowitz JA, Hernandez AF, O'Connor CM, et al. Assessment of dyspnea in acute decompensated heart failure: insights from ASCEND-HF (Acute Study of Clinical Effectiveness of Nesiritide in Decompensated Heart Failure) on the contributions of peak expiratory flow. *J Am Coll Cardiol.* 2012;59(16):1441-1448.
12. Gheorghiade M, Ruschitzka F. Beyond dyspnoea as an endpoint in acute heart failure trials. *Eur Heart J.* 2011;32(12):1442-1445.
13. Gheorghiade M, Konstam MA, Burnett JC, Jr., et al. Short-term clinical effects of tolvaptan, an oral vasopressin antagonist, in patients hospitalized for heart failure: the EVEREST Clinical Status Trials. *JAMA : the journal of the American Medical Association.* 2007;297(12):1332-1343.

14. McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. *JAMA : the journal of the American Medical Association.* 2014;312(13):1342-1343.
15. Ander DS, Aisiku IP, Ratcliff JJ, Todd KH, Gotsch K. Measuring the dyspnea of decompensated heart failure with a visual analog scale: how much improvement is meaningful? *Congestive heart failure.* 2004;10(4):188-191.
16. Gallagher EJ, Liebman M, Bijur PE. Prospective validation of clinically important changes in pain severity measured on a visual analog scale. *Annals of emergency medicine.* 2001;38(6):633-638.
17. Todd KH, Funk KG, Funk JP, Bonacci R. Clinical significance of reported changes in pain severity. *Annals of emergency medicine.* 1996;27(4):485-489.
18. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological).* 1995;57(1):289-300.
19. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health and quality of life outcomes.* 2006;4:70.
20. Cloutier MM, Schatz M, Castro M, et al. Asthma outcomes: composite scores of asthma control. *J Allergy Clin Immunol.* 2012;129(3 Suppl):S24-33.
21. Jones PW, Beeh KM, Chapman KR, Decramer M, Mahler DA, Wedzicha JA. Minimal clinically important differences in pharmacological trials. *Am J Respir Crit Care Med.* 2014;189(3):250-255.
22. Mannion AF, Balague F, Pellise F, Cedraschi C. Pain measurement in patients with low back pain. *Nat Clin Pract Rheumatol.* 2007;3(11):610-618.

Supplemental Table 1. Baseline Characteristics Grouped by Patient Reported 'Any Improvement', 'No Change' or 'Any Worsening' by 7-Point Likert

	Improved (n=398)	*No Change (n=68)*	*Worse (n=25)*	*3-group P-value*	*Minimally Improved vs. No Change P-value*
Age, mean ± SD	69.1 ± 14.3	62.6 ± 16.4	67.5 ± 13.9	0.0034	0.0008
Male gender, n (%)	226 (56.8%)	42 (61.8%)	12 (48.0%)	0.4662	0.5075
White race, n (%)	330 (83.1%) (n=397)	50 (73.5%)	21 (84.0%)	0.1766	0.0634
LVEF, mean ± SD	40.6 ± 16.9 (n=202)	36.6 ± 16.2 (n=43)	42.3 ± 17.8 (n=7)	0.3404	0.1550
HISTORY, n (%)					
Heart Failure	(n=386)			0.0139	0.0043
History of heart failure	231 (59.8%)	53 (77.9%)	16 (64.0%)	.	.
de novo heart failure	155 (40.2%)	15 (22.1%)	9 (36.0%)	.	.
Coronary artery disease	175 (44.0%)	26 (38.2%)	8 (32.0%)	0.3911	0.4275
Prior myocardial infarction	101 (25.4%)	18 (26.5%)	5 (20.0%)	0.8553	0.8806
Valvular disease	59 (14.8%)	14 (20.6%)	5 (20.0%)	0.3579	0.2772
Primary Cardiomyopathy	46 (11.6%)	18 (26.5%)	2 (8.0%)	0.0045	0.0021
Prior percutaneous coronary intervention (PCI)	42 (10.6%)	6 (8.8%)	4 (16.0%)	0.5328	0.8299
Prior coronary bypass grafts	33 (8.3%)	5 (7.4%)	1 (4.0%)	0.9443	1.0000
Stroke (CVA)	30 (7.5%)	4 (5.9%)	0 (0.0%)	0.4556	0.8028
Obesity	106 (26.6%)	12 (17.6%)	5 (20.0%)	0.2549	0.1322
Peripheral vascular disease	41 (10.3%)	1 (1.5%)	0 (0.0%)	0.0116	0.0191
Asthma/COPD	73 (18.3%)	8 (11.8%)	5 (20.0%)	0.3885	0.2265
Diabetes - Insulin Dependent	61 (15.3%)	13 (19.1%)	4 (16.0%)	0.7051	0.4721
Diabetes - Non-Insulin Dependent	82 (20.6%)	13 (19.1%)	4 (16.0%)	0.9243	0.8714
Renal insufficiency	107 (26.9%)	18 (26.5%)	3 (12.0%)	0.2779	1.0000
Anemia (HG<12 g/dl)	55 (13.8%)	8 (11.8%)	5 (20.0%)	0.5635	0.8477
Cancer history	25 (6.3%)	4 (5.9%)	2 (8.0%)	0.8064	1.0000
Hypertension	304 (76.4%)	49 (72.1%)	16 (64.0%)	0.2917	0.4462

VITAL SIGNS, mean ± SD					
Systolic Blood pressure (mm Hg)	146.2 ± 34.4	132.9 ± 32.5	132.8 ± 26.4	0.0032	0.0032
Diastolic Blood pressure (mm Hg)	82.7 ± 19.6 (n=397)	76.6 ± 20.8	74.4 ± 17.4	0.0127	0.0204
Heart rate (beats/min)	91.8 ± 24.3	89.4 ± 21.5	84.1 ± 13.3	0.2307	0.4430
Respiratory rate (breaths/min)	23.1 ± 6.3 (n=390)	21.9 ± 4.9 (n=66)	21.5 ± 6.3 (n=24)	0.1971	0.0841
Oxygen saturation (SpO2)	92.7 ± 5.8 (n=354)	95.6 ± 3.6 (n=66)	92.7 ± 8.0 (n=21)	0.0008	<0.0001
BASELINE MEDICATIONS, n(%)					
Beta-blockers	200 (50.3%)	41 (60.3%)	15 (60.0%)	0.2271	0.1487
ACE Inhibitors	181 (45.5%)	41 (60.3%)	10 (40.0%)	0.0610	0.0258
Angiotensin receptor blockers	52 (13.1%)	5 (7.4%)	4 (16.0%)	0.3455	0.2310
Statins	108 (27.1%)	20 (29.4%)	6 (24.0%)	0.8798	0.7688
Amlodipine/Felodipine (other dihydropyridine)	56 (14.1%)	8 (11.8%)	4 (16.0%)	0.8378	0.7059
Nitrates	84 (21.1%)	10 (14.7%)	3 (12.0%)	0.3598	0.2555
Aldosterone Antagonists	37 (9.3%)	15 (22.1%)	3 (12.0%)	0.0119	0.0054
Diuretics	252 (63.3%)	50 (73.5%)	17 (68.0%)	0.2548	0.1302
Cardiac Glycoside	75 (18.8%)	14 (20.6%)	2 (8.0%)	0.3727	0.7394
Aspirin	137 (34.4%)	31 (45.6%)	8 (32.0%)	0.1919	0.1004
Plavix	34 (8.5%)	9 (13.2%)	2 (8.0%)	0.4573	0.2539
Coumadin	64 (16.1%)	16 (23.5%)	2 (8.0%)	0.1708	0.1624
Other anti-coagulant	43 (10.8%)	4 (5.9%)	4 (16.0%)	0.3092	0.2776
Pacemaker	24 (6.0%)	4 (5.9%)	2 (8.0%)	0.7994	1.0000
Implantable cardioverter-defribrillators (AICD)	16 (4.0%)	5 (7.4%)	2 (8.0%)	0.2322	0.2113
Cardiac resynchronization therapy (CRT)	2 (0.5%)	0 (0.0%)	0 (0.0%)	1.0000	1.0000
LABORATORY VALUES, mean ± SD					
BNP	1374.9 ±1243.5 (n=94)	1031.2 ±1217.2 (n=26)	867.5 ± 873.6 (n=6)	0.3176	0.2128

Serum sodium (mmol/L)	137.9 ± 4.4 (n=386)	137.8 ± 5.3 (n=66)	137.8 ± 3.8	0.9743	0.8565
Troponin I >0.04 (ng/mL) label	5.8 ± 12.6 (n=159)	2.2 ± 8.2 (n=22)	7.5 ± 13.4 (n=6)	0.4120	0.0899
Creatinine	1.5 ± 1.3 (n=386)	1.7 ± 1.9 (n=67)	1.2 ± 0.5 (n=24)	0.3456	0.6014
BUN (mg/dL)	29.6 ± 25.3 (n=360)	29.4 ± 24.2 (n=62)	27.4 ± 19.9 (n=21)	0.9265	0.9549
TREATMENT IN THE ED, n (%)					
Loop diuretic	337 (84.9%) (n=397)	57 (86.4%) (n=66)	19 (76.0%)	0.4196	0.8535
Vasodilator	112 (30.0%) (n=373)	9 (14.1%) (n=64)	3 (13.0%) (n=23)	0.0078	0.0096
IV Inotrope and/or vasopressor	30 (8.1%) (n=370)	6 (9.2%) (n=65)	4 (16.7%) (n=24)	0.3071	0.8066
DYSPNEA SCALES					
5 point Likert (upright; n (%))	(n=394)	(n=67)		<.0001	<.0001
I am not SOB	16 (4.1%)	16 (23.9%)	2 (8.0%)	.	.
Mildly SOB	95 (24.1%)	23 (34.3%)	6 (24.0%)	.	.
Moderately SOB	134 (34.0%)	12 (17.9%)	7 (28.0%)	.	.
Severely SOB	83 (21.1%)	11 (16.4%)	8 (32.0%)	.	.
Very Severely SOB	66 (16.8%)	5 (7.5%)	2 (8.0%)	.	.
5 point Likert (supine, n(%))	(n=235)	(n=48)	(n=14)	<.0001	<.0001
I am not SOB	11 (4.7%)	13 (27.1%)	1 (7.1%)	.	.
Mildly SOB	43 (18.3%)	7 (14.6%)	7 (50.0%)	.	.
Moderately SOB	120 (51.1%)	25 (52.1%)	0 (0.0%)	.	.
Severely SOB	55 (23.4%)	2 (4.2%)	6 (42.9%)	.	.
Very Severely SOB	6 (2.6%)	1 (2.1%)	0 (0.0%)	.	.
Visual Analog Scale (upright), mean ± SD	5.8 ± 2.6	4.3 ± 3.0	5.6 ± 2.7	<.0001	<.0001
Visual Analog Scale (supine), mean ± SD	5.6 ± 2.2 (n=237)	4.0 ± 2.3 (n=49)	5.1 ± 3.0 (n=14)	<.0001	<.0001